PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/55 // 9/16

A1

(11) International Publication Number: WO 97/06262

(43) International Publication Date: 20 February 1997 (20.02.97)

(21) International Application Number: PCT/US96/12665

(22) International Filing Date: 1 August 1996 (01.08.96)

(30) Priority Data:

60/002,151	10 August 1995 (10.08.95)	US
08/567,507	5 December 1995 (05.12.95)	US

(71) Applicant: INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors: KENNEDY, Neil, F.; 6571 Jeremie Drive, San Jose, Ca 95120 (US). SEILHAMER, Jeffrey, J.; 12555 La Cresta, Los Altos Hills, CA 94022 (US). DELEGEANE, Angelo, M.; 30474 Oakmont Way, Hayward, CA 94544 (US). GUEGLER, Karl, J.; 1048 Oakland Avenue, Menlo Park, CA 94025 (US).

(74) Agent: BILLINGS, Lucy, J.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States: AT, AU, BR, CA, CH, CN, DE, DK, ES, FI, GB, IL, JP, KR, MX, NO, NZ, RU, SE, SG, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: NON-RECEPTOR TYPE HUMAN PROTEIN TYROSINE PHOSPHATASE

(57) Abstract

The present invention provides nucleotide and amino acid sequences that identify and encode a human homolog of rat PRL-1 derived from human mast cells. The present invention also provides for antisense molecules to the nucleotide sequences which encode HPRL, hybridization probes or oligonucleotides for the detection of HPRL-encoding nucleotide sequences, and a diagnostic test based on HPRL-encoding nucleic acid molecules. The present invention further provides for genetically engineered host cells for the expression of HPRL, biologically active HPRL, antibodies against HPRL, inhibitors and agonists of HPRL, and treatment methods comprising administration of compounds, such as antibodies, inhibitors or agonists.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

NON-RECEPTOR TYPE HUMAN PROTEIN TYROSINE PHOSPHATASE

TECHNICAL FIELD

The present invention is in the field of molecular biology; more particularly, the present invention describes the nucleic acid and amino acid sequences of a protein tyrosine phosphatase derived from human mast cells.

BACKGROUND ART

Protein phosphorylation and dephosphorylation control many signaling events related to cellular processes such as cell growth and differentiation, cell-to-cell contacts, the cell cycle and oncogenesis. Protein kinases increase protein phosphorylation, while protein phosphatases promote protein dephosphorylation to achieve a steady state level of phosphorylation. Most research has focused on understanding the structure and function of protein kinases in signaling events while neglecting that of protein phosphatases which play an equally important role.

There appear to be three distinct protein phosphatase gene families. They are 1) the acid/alkaline protein phosphatase gene family, whose role <u>in vivo</u> is not known, 2) the protein phosphatase or serine/threonine phosphatase gene family, and 3) the protein tyrosine phosphatase gene family.

The protein tyrosine phosphatase (PTPase) gene family consists of phosphatases that remove phosphate groups from protein tyrosine residues with high selectivity. One phosphorylated tyrosine residue may serve as a substrate, but another phosphotyrosine residue of the same protein may not. These phosphatases exist in a wide range of sizes and structural forms including transmembrane, receptor-like, and transmembrane forms. However, they all share homology within a region of 240 residues which defines a catalytic domain and contains a (I/V)HCXAGXXR(S/T)G consensus sequence near the C- terminus. To date about 30 distinct phosphatases have been identified.

The distinguishing features of Type I receptor-like phosphatases are the presence of a single transmembrane segment and one or two tandem consensus sequences within the cytoplasmic tail. Typically the extracellular sequences are widely divergent which probably reflects these protein's unique binding specificities. These transmembrane receptors may represent a new class of signal transduction molecules that activate intracellular signaling cascades by dephosphorylation of protein tryosyl residues. Type II receptor-like phosphatases contain one to three IgG-like domains near the amino terminus and 2 to 10 fibronectin type III-like repeats in tandem. Type III receptor-like phosphatases contain only tandem fibronectin type III-like repeats.

Both Type II and Type III receptor-like phosphatases may be involved in cell adhesion processes. Type IV receptor-like phosphatases are characterized by short extracellular sequences that are highly glycosylated and may be associated with additional polypeptides which may confer ligand binding capabilities.

Nonreceptor PTPases have single catalytic domains and noncatalytic domains of variable length positioned at either N- or C-terminus of the molecule. The noncatalytic regions share similarities to sequences found in other well characterized proteins and likely determine the subcellular localization of these proteins.

A subset of nonreceptor PTPases has been discovered that not only dephosphorylate phosphotyrosine, but also dephosphorylate phosphoserine/threonine residues in substrates. One such PTPase is the VH1 gene of vaccinia virus which encodes a 20-kd protein (Guan K., et al. (1991) Nature 350:359-62). VH-1 is able to dephosphorylate both phosphotyrosine and phosphoserine/threonine residues in artificial substrates such as casein and myelin basic protein. Another subset of nonreceptor PTPases possessing dual substrate specificity includes cdc25 which is involved in cell cycle regulation by dephosphorylating residues Thr^{14} and Tyr^{15} of cdc2. However sequence similarities in the catalytic domain between cdc25-like and VH1-like PTPases is low.

Yet another subset of nonreceptor PTPases includes the rat PRL-1 protein. It is a unique nuclear tyrosine phosphatase that controls cell growth. PRL-1 is a 20 kDa protein with the eight amino-acid consensus protein tyrosine phosphatase active site, HCXAGXXR, within its catalytic domain. Mutation of the active site cysteine residue abolishes this activity. PRL-1 has no homology to other PTPases outside the active site. The activity of the rat PRL-1 is regulated at the transcriptional level and additionally by specific intracellular localization, in this case primarily in the cell nucleus. PRL-1 is induced in mitogen-stimulated cells and expressed in regenerating liver, and its expression is elevated in a number of tumor cell lines. Stably transfected cells which overexpress PRL-1 demonstrate altered cellular growth and morphology and a transformed phenotype. It appears that PRL-1 is important in normal cellular growth control and may contribute to the tumorigenicity of some cancer cells (Diamond, RH et al. (1994) Mol. Cell. Biol. 14:3752-62).

Better and earlier methods for detecting tumors in human tissues or cells is necessary. These detection methods will lead to earlier treatment, less disability, and lower health care costs. Therefore, development of techniques using the polynucleotide or polypeptide of the subject invention is necessary for an early and accurate diagnosis

of abnormal cellular proliferation in humans.

DISCLOSURE OF THE INVENTION

The present invention discloses a novel, human tyrosine phosphatase (HPRL) characterized as having homology to rat PRL-1. The present invention also discloses a nucleotide sequence encoding HPRL that is shown in SEQ ID NO:1, between nucleotides 217 and 720. Alternatively, the nucleotide sequence may be any nucleotide sequence that is capable of expressing HPRL.

The present invention also discloses a polynucleotide sequence encoding HPRL and additional 5'-and 3'-sequences which may affect HPRL expression and transcript stability. Such control sequences are shown in SEQ ID NO:1 between nucleotides 0 and 216 and between nucleotides 721 and 1143.

Furthermore, the present invention provides cloning and expression vectors and host cells comprising polynucleotide sequences encoding HPRL.

The invention also provides a diagnostic test for conditions associated with altered levels of nucleic acid encoding HPRL and expression of HPRL, such as abnormal cellular proliferation typically associated with tumor growth.

A further aspect of the present invention includes HPRL antisense DNA which is employed to inhibit translation of HPRL mRNA.

Moreover, the present invention provides a method for the production and recovery of purified HPRL from host cells. The sequence for HPRL is shown in SEQ ID NO:2. Purified HPRL is used to produce antibodies against HPRL and to screen for inhibitors of HPRL activity. Antibodies and inhibitors of HPRL may be used to inhibit HPRL activity associated with abnormal cellular proliferation. Agonists of HPRL activity may be used to enhance HPRL activity in tissue regeneration.

The present invention also provides a method for treating conditions associated with abnormal cellular proliferation related to the altered expression of HPRL which comprises the administration of antibodies or inhibitors of HPRL activity as a pharmaceutical composition in an effective dosage.

BRIEF DESCRIPTION OF DRAWINGS

Figure 1 shows the amino acid alignment of HPRL with rat PRL-1. Alignments shown were produced using the multisequence alignment program of DNASTAR software (DNASTAR Inc, Madison WI).

Figure 2 shows a hydrophobicity plot for the amino acid sequence of HPRL using the hydrophobicity program of DNASTAR.

MODES FOR CARRYING OUT THE INVENTION

Definitions

The present invention discloses a novel, human tyrosine phosphatase (HPRL) as well as a unique polynucleotide sequence that encodes it (hprl) that has been isolated and purified from human mast cells. The term "isolated and purified" refers to hprl polynucleotide consisting of at least about 50% of the polynucleotide in a sample, preferably consisting of at least about 90% of the polynucleotide in a sample as assayed by standard techniques, such as gel electrophoresis.

As used herein, the lower case letters, hprl, refer to an mRNA, cDNA or nucleic acid sequence derived from the unique polynucleotide sequence. The nucleic acid sequence serves as an oligonucleotide or fragments.

An "oligonucleotide" is a portion of a DNA sequence which has a sufficient number of bases to be used as an oligomer, probe, amplimer or primer in a polymerase chain reaction (PCR). Oligonucleotides are prepared from genomic or cDNA sequences and are used to amplify, confirm, or reveal the presence of an identical or similar DNA or RNA in a particular cell or tissue. Oligonucleotides comprise portions of a DNA sequence having at least about 10 nucleotides and as many as about 50 nucleotides, preferably about 15 to 30 nucleotides.

A "fragment" of a polynucleotide comprises a part of the polynucleotide sequence having fewer nucleotides than about 1.2 kilobase pairs (kb), preferably between about 1 kb and 0.1kb, which can be used as a probe, can be therapeutically active (i.e. as antisense DNA), or can be used to synthesize a polypeptide fragment. Probes are useful in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are highly specific, but slow to hybridize. Shorter length probes are quicker to hybridize, but must be carefully designed to have specificity. Single- or double-stranded probes may be either chemically synthesized or obtained and/or modified from naturally occurring or recombinant sequences. Such probes may be labelled with reporter molecules using nick translation, Klenow fill-in reaction, PCR and other methods well known in the art. After pretesting to optimize reaction conditions and to eliminate false positives, nucleic acid probes may be used in Southern, northern or in situ hybridizations to determine whether DNA or RNA encoding the protein is present in a biological sample, cell type, tissue, organ or organism.

"Reporter" or "label" molecules are chemical moieties used for labelling a nucleic or an amino acid sequence. They include, but are not limited to, radionuclides, enzymes, fluorescent, chemiluminescent, and chromogenic agents. Reporter molecules associate with particular

nucleic acid or amino acid sequence for determining its presence and its quantity in a sample. Additionally, reporter molecules can be used therapeutically.

A "recombinant nucleotide variant" is a nucleotide sequence that has been altered from that which occurs naturally. Various codon substitutions, such as the silent changes which produce specific restriction sites or codon usage-specific mutations, may be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic host system, respectively. In particular, the sequence encodes a polypeptide which may be synthesized by making use of the "redundancy" in the genetic code. Recombinant nucleotide variants also include nucleotide sequences with the necessary nucleic acid substitutions, insertions and/or deletions to produce recombinant polypeptide variants (described below).

"Linkers" are synthesized palindromic oligomers which create internal restriction endonuclease sites for ease of cloning the genetic material of choice into various vectors.

As used herein, the upper case letters, HPRL, refer to the polypeptide, oligopeptide or fragment derived therefrom. The term "isolated and purified HPRL polypeptide" refers to a HPRL polypeptide sample consisting of greater than about 50% of the isolated and purified polypeptide, preferably greater than about 90%, in a sample as identified by standard laboratory techniques, such as gel electrophoresis or chromatographic methods.

The term "chimeric molecule", as used herein, refers to polynucleotides or polypeptides which are created by combining one or more nucleotide sequences of this invention with additional nucleic acid sequence(s). Such combined sequences may be introduced into an appropriate vector for expression of a chimeric polypeptide. Examples of useful chimeric polypeptides are those with changes in cellular location, distribution, ligand-binding affinities, interchain affinities, degradation/turnover rate, signalling, etc.

"Active" refers to those forms, fragments, or domains of a polypeptide sequence which display biologic and/or immunogenic properties characteristic of the naturally occurring polypeptide.

"Naturally occurring HPRL" refers to a polypeptide produced by cells which have not been genetically engineered or which have been genetically engineered to produce the same sequence as that naturally produced. Specifically contemplated are various polypeptides which arise from post-translational modifications. Such modifications of the polypeptide include but are not limited to acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation.

"Derivative" refers to a polypeptide which has been chemically modified by such techniques as ubiquitination, labelling (see above), pegylation (derivatization with polyethylene glycol), and chemical insertion, deletion and substitution of amino acids. It comprises substitutions of amino acids such as ornithine which do not normally occur in human proteins.

"Recombinant polypeptide variant" refers to any polypeptide which differs from naturally occurring HPRL by amino acid insertions, deletions and/or substitutions, created using recombinant DNA techniques. Guidance in determining which amino acid residues may be replaced, added or deleted without abolishing characteristics of interest may be found by comparing the sequence of HPRL with that of related polypeptides and minimizing the number of amino acid changes in highly conserved regions.

Amino acid "substitutions" are defined as one for one amino acid replacements. They are conservative in nature when the substituted amino acid has similar structural and/or chemical properties. Examples of conservative replacements are substitution of a leucine with an isoleucine or valine, an aspartate with a glutamate, or a threonine with a serine. Examples of nonconservative replacements are substitutions of a glutamate with a glutamine, or an aspartate with an asparagine.

Amino acid "insertions" or "deletions" are additions or eliminations of amino acids from polypeptides. They typically fall in the range of about 1 to 5 amino acids. The variation allowed in a particular amino acid sequence may be experimentally determined by producing the peptide synthetically or by systematically making insertions, deletions, or substitutions of nucleotides in the hprl sequence using recombinant DNA techniques.

A "signal or leader sequence" is a short amino acid sequence which directs, or can be used to direct, the polypeptide through the membrane of a cell. Such a sequence may be naturally present on the polypeptides of the present invention. Also, signal sequences can be provided from heterologous sources and added to the polypeptide by recombinant DNA techniques.

An "oligopeptide" is a short stretch of amino acid residues and may be expressed from an oligonucleotide. It may be considerably shorter than or the same length as a "fragment," "portion," or "segment" of a polypeptide. An oligopeptide comprises a stretch of amino acid residues of at least about 5 amino acids and often about 17 or more amino acids, typically at least about 9 to 13 amino acids, and of sufficient length to display biologic and/or immunogenic activity.

A "standard" is the measurement of a compound at known

concentrations used for comparing samples with unknown concentrations of the same compound. Preferably, the standard is based on a statistically appropriate number of samples. Furthermore, the standard is used for comparing samples with unknown concentrations of a related compound. When comparing samples with unknown concentrations of a related compound, the standard is corrected by a factor that reflects differences, including biological and physical differences, between the compound and the related compound. Standards are used as a basis of comparison when performing diagnostic assays for monitoring clinical trials, or following patient treatment profiles.

"Animal" as used herein includes human, domestic (cats, dogs, etc.), agricultural (cows, fish, horses, sheep, chickens, etc) and test species (amphibian, rodent, simian, etc).

Since the list of technical and scientific terms cannot be all encompassing, any undefined terms shall be construed to have the same meaning as is commonly understood by one of ordinary skill in the art to which this invention belongs. Furthermore, the singular forms "a", "an" and "the" include plural referents unless the context clearly dictates otherwise. For example, reference to a "high-fidelity enzyme" includes mixtures of such enzymes and any other enzymes fitting the stated criteria, and reference to "the method" includes reference to one or more methods for doing the same thing, which will be known to those skilled in the art or will become known to them upon reading this specification.

Before the present sequences, variants, formulations and methods for making and using the invention are described, it is to be understood that the invention is not to be limited only to the particular sequences, variants, formulations or methods described. The sequences, variants, formulations and methodologies may vary, and the terminology used herein is for the purpose of describing particular embodiments. The terminology and definitions are not intended to be limiting since the scope of protection will ultimately depend upon the claims.

Detailed Description of the Invention

The present invention discloses a novel human tyrosine phosphatase (SEQ ID NO:2) characterized as having homology to rat PRL-1 (HPRL). HPRL is a protein tyrosine phosphatase, ie it recognizes and dephosphorylates tyrosine residues of a protein substrate. HPRL likely functions in the control of cellular proliferation, particularly cellular proliferation in tumor cell lines, in mitogen-stimulated cells and during tissue regeneration. Overexpression of HPRL therefore may contribute to the tumorigenicity of some cancer

cells, such as that of the mast cell leukemic cells. The present invention also discloses a unique polynucleotide sequence encoding HPRL, designated herein as hprl. The hprl polynucleotide sequence is included in SEQ ID NO:1 at nucleotides 217 to 720. SEQ ID NO:1 also includes 5'-and 3'-nucleotide sequences of the hprl coding sequence that function to control the stability of the mRNA and the expression of HPRL polypeptide. The 5'-sequence includes nucleotides 0 to 216. The 3'-sequence includes nucleotides 721 to 1144.

The polynucleotide sequence encoding HPRL was first identified in Incyte Clone 3752 which was isolated from a human mast cell cDNA library prepared from a human mast cell-1 cell line (HMC-1) established from the peripheral blood of an individual having mast cell leukemia. Therefore, this cDNA library may contain nucleic acid encoding novel polypeptides that control cellular proliferation.

Mast cells play an important role in promoting various immune responses and nonspecific inflammatory reactions. Their major function is secretion via degranulation discharging granule contents into the extracellular environment. These cells are also capable of phagocytosis and pinocytosis, and may serve a storage function for substances like biogenic amines and fatty acids.

The purified nucleic acid sequence for hprl has numerous applications in techniques known to those skilled in the art of molecular biology. These applications include its use as a hybridization probe, for chromosome and gene mapping, in PCR technologies, in the production of sense or antisense nucleic acids, in screening for new therapeutic molecules, etc. These examples are well known and are not intended to be limiting. Furthermore, hprl may be used in molecular biology techniques that have not yet been developed, provided the new techniques rely on properties of nucleotide sequences that are currently known such as the triplet genetic code, specific base pair interactions, etc.

As a result of the degeneracy of the genetic code, a multitude of hprl-encoding nucleotide sequences may be produced and some of these will bear only minimal homology to the endogenous sequence of any known and naturally occurring hprl. This invention has specifically contemplated each and every possible variation of nucleotide sequence that could be made by selecting combinations based on possible codon choices.

Although the hprl nucleotide sequence and its derivatives or variants are preferably capable of identifying the nucleotide sequence of the naturally occurring hprl under optimized conditions, it may be advantageous to produce hprl nucleotide sequences possessing a substantially different codon usage. Codons can be selected to

increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic expression host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the hprl nucleotide sequence without altering the encoded amino acid sequence include the production of RNA transcripts having more desirable properties, such as a longer half-life, than transcripts produced from the naturally occurring sequence.

Nucleotide sequences for hprl may be joined to a variety of other nucleotide sequences by means of well established recombinant DNA techniques (Sambrook J et al (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor NY; or Ausubel FM et al (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York City). Useful sequences for joining to hprl include an assortment of cloning vectors such as plasmids, cosmids, lambda phage derivatives, phagemids, and the like. Vectors of interest include vectors for replication, expression, probe generation, sequencing, and the like. In general, vectors of interest may contain an origin of replication functional in at least one organism, convenient restriction endonuclease sensitive sites, and selectable markers for one or more host cell systems.

PCR as described in U.S. Patent Nos. 4,683,195; 4,800,195; and 4,965,188 provides additional uses for oligonucleotides based upon the hprl nucleotide sequence. Such oligomers are generally chemically synthesized, but they may be of recombinant origin or a mixture of both. Oligomers may comprise two nucleotide sequences employed under optimized conditions for tissue specific identification or diagnostic use. The same two oligomers, nested sets of oligomers, or even a degenerate pool of oligomers may be employed under less stringent conditions for identification and/or quantitation of closely related DNA or RNA sequences.

Full length genes may be cloned utilizing partial nucleotide sequence and various methods known in the art. "Restriction-site PCR" is a direct method (Gobinda et al (1993) PCR Methods Applic. 2:318-22) which uses universal primers to retrieve unknown sequence adjacent to a known locus. First, genomic DNA is amplified in the presence of high fidelity enzymes, a primer adjacent to linker, and a primer specific adjacent to the known region. The amplified sequences are subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR (Triglia T et al (1988) Nucleic Acids Res. 16:8186)

is the first method to report successful acquisition of unknown sequences starting with primers based on a known region. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template. Divergent primers are designed to prime outward from the known region and multiple rounds of restriction enzyme digestions and ligations are necessary.

Capture PCR (Lagerstrom M et al (1991) PCR Methods Applic. 1:111-19) is a method for PCR amplification of DNA fragments adjacent to a known human sequence in a yeast artificial chromosome (YAC). Capture PCR also requires multiple restriction enzyme digestions and ligations to place an engineered double-stranded sequence into an unknown portion of the DNA molecule before PCR. This method allows the restriction and ligation reactions to be carried out simultaneously and further requires extension, immobilization, two rounds of PCR and purification prior to sequencing.

Parker JD et al (1991; Nucleic Acids Res. 19:3055-60) teach walking PCR, a method for targeted gene walking which permits the retrieval of unknown sequence. For use of this method, PromoterFinder™, a new kit available from Clontech (Palo Alto CA), uses PCR and primers derived from p53 to walk in genomic DNA. Nested primers and special PromoterFinder libraries are used to detect upstream sequences such as promoters and regulatory elements. This process avoids the need to screen libraries and is useful in finding intron/exon junctions. A new method which employs XL-PCR (Perkin-Elmer, Foster City, CA) amplifies and extends partial nucleotide sequence into long pieces of DNA. This method was developed to allow a single researcher to process multiple genes (up to 20 or more) at a time and to obtain an extended (possibly full-length) sequence within 6-10 days. This new method replaces methods which use labelled probes to screen plasmid libraries and allow one researcher to process only about 3-5 genes in 14-40 days. A description of the method is provided in co-pending Application Serial No. 08/487,112 filed June 7, 1995, entitled "Improved Method for Obtaining Full Length cDNA Sequences" which is herein incorporated by reference.

In the first step, which can be performed in about two days, any two of a plurality of primers are designed and synthesized based on a known partial sequence. In step 2, which takes about six to eight hours, the sequence is extended by PCR amplification of a selected library. Steps 3 and 4, which take about one day, are purification of the amplified cDNA and its ligation into an appropriate vector. Step 5, which takes about one day, involves transforming and growing up host bacteria. In step 6, which takes approximately five hours, PCR

is used to screen bacterial clones for extended sequence. The final steps, which take about one day, involve the preparation and sequencing of selected clones.

If the full length cDNA has not been obtained, the entire procedure is repeated using either the original library or some other preferred library. The preferred library may be one that has been size-selected to include only larger cDNAs or may consist of single or combined commercially available libraries, eg. lung, liver, heart and brain from Gibco/BRL (Gaithersburg MD). The cDNA library may have been prepared with oligo (dT) or random priming. Random primed libraries are preferred in that they will contain more sequences which contain 5' ends of genes. A randomly primed library may be particularly useful if an oligo (dT) library does not yield a complete gene. It must be noted that the larger and more complex the protein, the less likely it is that the complete gene will be found in a single plasmid.

A new method for analyzing either the size or the nucleotide sequence of PCR products is capillary electrophoresis. Systems for rapid sequencing are available from Perkin Elmer (Foster City CA), Beckman Instruments (Fullerton CA), and other companies. Capillary sequencing employs flowable polymers for electrophoretic separation, four different fluorescent dyes (one for each nucleotide) which are laser activated, and detection of the emitted wavelengths by a charge coupled devise camera. Output/light intensity is converted to electrical signal using appropriate software (eg. Genotyper™ and Sequence Navigator™ from Perkin Elmer) and the entire process from loading of samples to computer analysis and electronic data display is computer controlled. Capillary electrophoresis provides greater resolution and is many times faster than standard gel based procedures. It is particularly suited to the sequencing of small pieces of DNA which might be present in limited amounts in a particular sample. The reproducible sequencing of up to 350 bp of M13 phage DNA in 30 min has been reported (Ruiz-Martinez MC et al (1993) Anal Chem 65:2851-8).

Another aspect of the subject invention is to provide for a hprl hybridization probe which is capable of hybridizing with naturally occurring nucleotide sequences for hprl. The stringency of the hybridization conditions will determine whether the probe identifies only the native hprl sequence or a sequence of closely related molecules. If a degenerate hprl sequence of the subject invention is used for the detection of related sequences, it should preferably contain at least 50% of the nucleotides of the sequence presented herein. Hybridization probes may be derived from the nucleotide

sequence of the SEQ ID NO:1, or from surrounding or included genomic sequences comprising untranslated regions such as promoters, enhancers and introns. Such hybridization probes may be labelled with appropriate reporter molecules.

Means for producing specific hybridization probes for hprl include oligolabelling, nick translation, end-labelling or PCR amplification using a labelled nucleotide. Alternatively, the cDNA sequence may be cloned into a vector for the production of mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3 or SP6 and labelled nucleotides. A number of companies which develop molecular biology products (such as Pharmacia Biotech, Piscataway NJ; Promega, Madison WI; USB, Cleveland OH, etc.) supply commercial kits and protocols for these various procedures.

It is also possible to produce a DNA sequence, or portions thereof, entirely by synthetic chemistry. Sometimes the source of information for producing this sequence comes from a known homologous sequence from a closely related organism. After synthesis, the nucleic acid sequence can be used alone or joined with other sequence(s) and inserted into one of the many available DNA vectors and their respective host cells using techniques well known in the art. Moreover, synthetic chemistry may be used to introduce specific mutations into the nucleotide sequence. Alternatively, a portion of sequence in which a mutation is desired can be synthesized and recombined with a portion of an existing genomic or recombinant sequence.

The nucleotide sequence for hprl can be used in an assay to detect or diagnose conditions associated with the altered hprl genomic sequences or altered levels of hprl mRNA in mast cells or other tissues. Detrimental conditions relate to excessive cellular proliferation likely leading to a tumor growth. Beneficial conditions relate to tissue regeneration. The cDNA can be labeled by methods known in the art and added to a fluid, cell or tissue sample from a patient under hybridizing conditions. After an incubation period, the sample is washed with a compatible fluid which optionally contains a reporter molecule. After the compatible fluid is rinsed off, the reporter molecule is quantitated and compared with a standard as previously defined. If hprl mRNA levels are significantly different from normal levels, the assay may indicate abnormal cellular proliferation.

This same assay, combining a sample with the nucleotide sequence, is applicable in evaluating the efficacy of a particular

therapeutic treatment. It may be used in animal studies, in clinical trials, or in monitoring the treatment of an individual patient. First, standard expression must be established for use as a basis of comparison. Second, samples from the animals or patients affected by the disease are combined with the nucleotide sequence to evaluate the deviation from the standard or normal profile. Third, an existing therapeutic agent is administered, and a treatment profile is generated. The assay is evaluated to determine whether the profile progresses toward or returns to the standard pattern. Successive treatment profiles may be used to show the efficacy of treatment over a period of several days or several months.

The hprl nucleotide sequence can also be used to generate probes for mapping the native gene. The sequence may be mapped to a particular chromosome or to a specific region of the chromosome using well known techniques. These include in situ hybridization to chromosomal spreads (Verma et al (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York City), flow-sorted chromosomal preparations, or artificial chromosome constructions such as YACs, bacterial artificial chromosomes (BACs), bacterial P1 constructions or single chromosome cDNA libraries.

In situ hybridization of chromosomal preparations and physical mapping techniques such as linkage analysis using established chromosomal markers are invaluable in extending genetic maps. Examples of genetic maps can be found in the 1994 Genome Issue of Science (265:1981f). Often the placement of a gene on the chromosome of another mammalian species may reveal associated markers even if the number or arm of a particular human chromosome is not known. New partial nucleotide sequences can be assigned to chromosomal arms, or parts thereof, by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once a disease or syndrome, such as ataxia telangiectasia (AT), has been crudely localized by genetic linkage to a particular genomic region, for example, AT to 11q22-23 (Gatti et al (1988) Nature 336:577-580), any sequences mapping to that area may represent genes for further investigation. The nucleotide sequences of the subject invention may also be used to detect differences in the chromosomal location of nucleotide sequences due to translocation, inversion, etc. between normal and carrier or affected individuals.

The hprl nucleotide sequence may be used to produce an amino acid sequence using well known methods of recombinant DNA technology (Goeddel (1990) Gene Expression Technology, Methods in Enzymology, Vol 185, Academic Press, San Diego CA). In some cases, genetically

engineered HPRL molecules that have a higher activity than naturally-occuring HPRL are of interest, such as for enhancing tissue regeneration. Alternatively, of interest are genetically engineered HPRL molecules that have a lower activity than naturally-occuring HPRL for controlling abnormal cellular proliferation. The HPRL amino acid sequence or a fragment thereof may be expressed in a variety of host cells, either prokaryotic or eukaryotic. Host cells may be from the same species from which the nucleotide sequence was derived or from a different species. Advantages of producing an amino acid sequence or peptide by recombinant DNA technology include the availability of simplified purification procedures to generate large quantities of polypeptide or fragments thereof.

Cells transformed with hprl may be cultured under conditions suitable for the expression and recovery of the polypeptide from cell culture. The polypeptide produced by a recombinant cell may be secreted or may be contained intracellularly depending on the sequence itself and/or the vector used. In general, it is more convenient to prepare recombinant polypeptides in secreted form, and this is accomplished by ligating hprl to a recombinant nucleotide sequence which directs its movement through a particular prokaryotic or eukaryotic cell membrane. Other chimeric constructions may join hprl to nucleotide sequence encoding a polypeptide domain which will facilitate protein purification (Kroll DJ et al (1993) DNA Cell Biol 12:441-53).

Direct peptide synthesis using solid-phase techniques (Stewart et al (1969) Solid-Phase Peptide Synthesis, WH Freeman Co, San Francisco CA; Merrifield J (1963) J Am Chem Soc 85:2149-2154) is an alternative to recombinant or chimeric peptide production. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Foster City CA) in accordance with the instructions provided by the manufacturer. Additionally the HPRL sequence or any fragment thereof may be mutated during direct synthesis and, if desired, combined using chemical methods with other amino acid sequences.

Another aspect of the subject invention is the use of HPRL specific antibodies to detect increased expression of HPRL, in particular, related to abnormal cellular proliferation including that related to tumor growth. Alternatively, these antibodies may be employed to monitor tissue regeneration by observing the increased expression of HPRL in certain tissues, such as liver.

HPRL used to induce specific antibodies may have an amino acid sequence consisting of at least five amino acids and preferably at least 10 amino acids, up to and including the entire HPRL sequence.

Short stretches of amino acid sequence up to the full length sequence of HPRL may be fused with those of another protein such as keyhole limpet hemocyanin, and the chimeric peptide used for antibody production.

Antibodies specific for HPRL may be produced by inoculation of an appropriate animal with an HPRL antigen as described above. An antibody is specific for HPRL if it binds to at least part of the polypeptide. Antibody production includes not only the stimulation of an immune response by injection into animals, but also analogous processes such as the production of synthetic antibodies, the screening of recombinant immunoglobulin libraries for specific-binding molecules (Orlandi R et al (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837, or Huse WD et al (1989) Science 256:1275-1281), or the *in vitro* stimulation of lymphocyte populations.

Current technology (Winter G and Milstein C (1991) Nature 349:293-299) provides for a number of highly specific binding reagents based on the principles of antibody formation. These techniques may be adapted to produce molecules which specifically bind HPRL. Antibodies or other appropriate molecules generated against a specific immunogenic peptide fragment or oligopeptide can be used in Western analysis, enzyme-linked immunosorbent assays (ELISA) or similar tests to establish the presence of or to quantitate amounts of HPRL active in normal, diseased, or therapeutically treated cells or tissues. Variations on any procedure known in the art for the measurement of HPRL can be used in the practice of the instant invention. Additionally, these procedures include but are not limited to competitive and non-competitive assay systems using techniques such as radioimmunoassays, sandwich immunoassays, agglutination assays, complement fixation assays, immunoradiometric assays, fluorescent immunoassays, protein A immunoassays, and immunoelectrophoresis assays.

Various procedures known in the art may be used for the production of antibody. Various animal species may be immunized with the HPRL, or a fragment thereof, and various adjuvants may be used to increase the immunological response, depending on the animal species, and including but are not limited to Freund's adjuvant (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, polyanions, peptides, oil emulsion, keyhole limpet hemocyanin, dinitrophenol, or liposomes.

A monoclonal antibody can be prepared by using any technique which provides for the production of antibody molecules by continuous cell lines in cultures. These include but are not limited to the hybridoma technique originally described by Kohler and Milstein (1975)

Nature 256:495-497), the more recent human B cell hybridoma technique (Kozbor et al (1983) Immunol Today 4:72) and EBV-hybridoma technique (Cole et al (1985) Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc).

In one embodiment, the monoclonal antibodies may be human monoclonal antibodies, chimeric human-mouse (or other species), or humanized antibodies. Human monoclonals may be made by any of numerous techniques known in the art (Kozbor). Chimeric antibody molecules may be prepared containing a mouse or other species) antigen-binding domain with human constant regions (Morrison et al. (1984) Proc. Natl. Acad. Sci. USA 81:6852). Humanized antibodies may be recombinantly prepared such that only the hypervariable domains are non-human sequences.

Kits for carrying out the assays are also within the scope of the invention. A kit for detecting HPRL can be prepared for routine use. Such a kit would include wells to receive a sample from a patient. Each well contains a reagent capable of binding HPRL selectively, particularly a monoclonal antibody against purified HPRL or a fragment thereof. The kits naturally include both negative and positive controls. The negative control is a sample lacking HPRL. The positive controls are samples with known concentrations of HPRL. In addition, the kits may contain a detection antibody, washing solutions, and a substrate used for generating a colored solution from the detection solution.

Administration of HPRL, fragments, derivatives or agonists may enhance tissue-regeneration, whereas administration of antibodies against HPRL or HPRL inhibitors may inhibit cellular proliferation of a tumor.

Bioactive compositions comprising HPRL, its antibodies, inhibitors or agonists may be administered in a suitable therapeutic dose determined by any of several methodologies including clinical studies on mammalian species to determine maximal tolerable dose and on normal human subjects to determine safe dose. Additionally, the bioactive agent may be complexed with a variety of well established compounds or compositions which enhance stability or pharmacological properties such as half-life. It is contemplated that the therapeutic, bioactive composition may be delivered by intravenous infusion into the bloodstream or any other effective means which could be used for treating conditions related to the altered expression of HPRL.

The examples below are provided to illustrate the subject invention. These examples are provided by way of illustration and are not included for the purpose of limiting the invention.

INDUSTRIAL APPLICABILITY

I cDNA Library Construction

The human mast cell-1 (HMC-1) cell line was established from the peripheral blood of a Mayo clinic patient with mast cell leukemia (Leukemia Research 12:345-355). The cultured cells looked similar to immature cloned murine mast cells, contained histamine, and stained positively for chloroacetate esterase, amino caproate esterase and tryptase. The HMC-1 cells, however, lacked the ability to synthesize normal IgE receptors.

Cultured HMC-1 cells were lysed immediately in buffer containing guanidinium isothiocyanate. Lysis was followed by several phenol chloroform extractions and ethanol precipitation. Poly(A+)mRNA was isolated using biotinylated oligo d(T) primer and streptavidin coupled to a paramagnetic particle (Promega Corporation, Madison WI) and sent to Stratagene (Stratagene, La Jolla, CA).

The human mast cell HMC-1 cDNA library was custom constructed by Stratagene essentially as described. Complementary DNA (cDNA) synthesis was primed separately with either oligo dT or random hexamers for the enzymatic synthesis of double stranded cDNA. Synthetic adaptor oligonucleotides were ligated onto cDNA ends enabling its insertion into the lambda vector. The HMC-1 cDNA library was constructed using the Uni-ZAP™ vector system (Stratagene), allowing high efficiency unidirectional (sense orientation) lambda library construction and the convenience of a plasmid system with blue/white color selection to detect clones with cDNA insertions.

The HMC-1 cDNA library can be screened with either DNA probes or antibody probes and the pBluescript® phagemid (Stratagene) can be rapidly excised in vivo. The phagemid allows the use of a plasmid system for easy insert characterization, sequencing, site-directed mutagenesis, the creation of unidirectional deletions and expression of fusion polypeptides. The custom-constructed library phage particles were infected into E. Coli host strain XL1-Blue® (Stratagene), which has a high transformation efficiency, increasing the probability of obtaining rare, under-represented clones in the cDNA library. Alternative unidirectional vectors include but are not limited to pcDNAI (Invitrogen, San Diego CA) and pSHlox-1 (Novagen, Madison WI).

II Isolation of cDNA Clones

The phagemid forms of individual cDNA clones were obtained by the in vivo excision process, in which the host bacterial strain was coinfected with both the λ library phage and an f1 helper phage. Polypeptides or enzymes derived from both the library-containing phage and the helper phage nicked the λ DNA, initiated new DNA synthesis

from defined sequences on the λ target DNA and created a smaller, single stranded circular phagemid DNA molecule that included all DNA sequences of the pBLUESCRIPT® plasmid and the cDNA insert. The phagemid DNA was secreted from the cells and purified, then used to reinfect fresh host cells, where the double stranded phagemid DNA was produced. Because the phagemid carries the gene for β-lactamase, the newly transformed bacteria were selected on medium containing ampicillin.

Phagemid DNA was purified using the MAGIC MINIPREPS™ DNA Purification System (Promega catalogue #A7100, Promega Corporation, Madison WI). This small-scale process provides a simple and reliable method for lysing the bacterial cells and rapidly isolating purified phagemid DNA using a proprietary DNA-binding resin. The DNA was eluted from the purification resin already prepared for DNA sequencing and other analytical manipulations. Phagemid DNA was also purified using the QIAWELL-8 Plasmid Purification System from QIAGEN®, QIAWELL PLUS and QIAWELL ULTRA DNA Purification System (QIAGEN Inc., Chatsworth, CA). This product line provides a convenient, rapid and reliable high-throughput method for lysing the bacterial cells and isolating highly purified phagemid DNA using QIAGEN anion-exchange resin particles with EMPORE™ membrane technology from 3M in a multiwell format.

III Sequencing of cDNA Clones

The cDNA inserts from random isolates of the human mast cell library were sequenced in part. Methods for DNA sequencing are well known in the art. Conventional enzymatic methods employed DNA polymerase Klenow fragment, SEQUENASE® (US Biochemical Corp, Cleveland, OH) or Taq polymerase to extend DNA chains from an oligonucleotide primer annealed to the DNA template of interest. Methods have been developed for the use of both single- and double-stranded templates. The chain termination reaction products were electrophoresed on urea-acrylamide gels and detected either by autoradiography (for radionuclide-labeled precursors) or by fluorescence (for fluorescent-labeled precursors). Recent improvements in mechanized reaction preparation, sequencing and analysis using the fluorescent detection method have permitted expansion in the number of sequences that can be determined per day (using machines such as the Catalyst 800 and the Applied Biosystems 377 or 373 DNA sequencer). Alternatively, cDNA inserts may be sequenced using a Hamilton Micro Lab 2200 (Hamilton, Reno NV) in combination with four Peltier Thermal Cyclers (PTC200 from MJ Research, Watertown MA) along with Applied Biosystems 377 or 373 DNA Sequencing System.

IV Identification and Full Length Sequencing of the Genes

The nucleotide sequence for the entire hprl coding region claimed in this invention is shown in SEQ ID NO:1 from nucleotides 217 to 720. The hprl coding region identified in Incyte Clone 3752 was extended to the full length hprl coding sequence using a modified XL-PCR (Perkin Elmer) procedure. Primers were designed based on known sequence; one primer was synthesized to initiate extension in the antisense direction (XLR) and the other to extend sequence in the sense direction (XLF). The sequences of these primers and their location are as follows: XLR (nucleotides 487 to 509 in SEQ ID NO:1) and XLF (nucleotides 553 to 578 in SEQ ID NO:1). The primers allowed the sequence to be extended "outward" generating amplicons containing new, unknown nucleotide sequence for the gene of interest. The primers were designed using Oligo 4.0 (National Biosciences Inc, Plymouth MN) to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68°-72° C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations were avoided.

The HMC-1 cDNA library was used as a template, and XLR and XLF primers were used to extend and amplify the coding sequence in Incyte Clone 3752. By following the instructions for the XL-PCR kit, the enzymes provided high fidelity in the amplification. Beginning with 25 pMol of each primer and the recommended concentrations of all other components of the kit, PCR was performed using the MJ PTC200 (MJ Research, Watertown MA) and the following parameters:

Step 1	94° C for 60 sec (initial denaturation)
Step 2	94° C for 15 sec
Step 3	65° C for 1 min
Step 4	68° C for 7 min
Step 5	Repeat step 2-4 for 15 additional cycles
Step 6	94° C for 15 sec
Step 7	65° C for 1 min
Step 8	68° C for 7 min + 15 sec/cycle
Step 9	Repeat step 6-8 for 11 additional cycles
Step 10	72° C for 8 min
Step 11	4° C (and holding)

At the end of 28 cycles, 50 µl of the reaction mix was removed; and the remaining reaction mix was run for an additional 10 cycles as outlined below:

Step 1	94° C for 15 sec
Step 2	65° C for 1 min
Step 3	68° C for (10 min + 15 sec)/cycle
Step 4	Repeat step 1-3 for 9 additional cycles
Step 5	72° C for 10 min

A 5-10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a low concentration, about 0.6-0.8%, agarose mini-gel to determine which reactions were successful in extending the sequence. Although all extensions potentially contain a full length

gene, some of the largest products or bands were selected and cut out of the gel. Further purification involved using a commercial gel extraction method such as QIAQuick™ (QIAGEN Inc, Chatsworth CA). After recovery of the DNA, Klenow enzyme was used to trim single stranded, nucleotide overhangs creating blunt ends which facilitated religation and cloning.

After ethanol precipitation, the products were redissolved in 13 μl of ligation buffer. Then, 1μl T4-DNA ligase (15 units) and 1μl T4 polynucleotide kinase were added, and the mixture was incubated at room temperature for 2-3 hours or overnight at 16° C. Competent E. coli cells (in 40 μl of appropriate media) were transformed with 3 μl of ligation mixture and cultured in 80 μl of SOC medium (Sambrook J et al, supra). After incubation for one hour at 37° C, the whole transformation mixture was plated on Luria Bertani (LB)-agar (Sambrook J et al, supra) containing carbenicillin at 25 mg/L. The following day, 12 colonies were randomly picked from each plate and cultured in 150 μl of liquid LB/carbenicillin medium placed in an individual well of an appropriate, commercially-available, sterile 96-well microtiter plate. The following day, 5 μl of each overnight culture was transferred into a non-sterile 96-well plate and after dilution 1:10 with water, 5 μl of each sample was transferred into a PCR array.

For PCR amplification, 15 μl of PCR mix (1.33x containing 0.75 units of Taq polymerase, a vector primer and one or both of the gene specific primers used for the extension reaction) were added to each well. Amplification was performed using the following conditions:

Step 1	94° C for 60 sec
Step 2	94° C for 20 sec
Step 3	55° C for 30 sec
Step 4	72° C for 90 sec
Step 5	Repeat steps 2-4 for an additional 29 cycles
Step 6	72° C for 180 sec
Step 7	4° C (and holding)

Aliquots of the PCR reactions were run on agarose gels together with molecular weight markers. The sizes of the PCR products were compared to the original partial cDNAs, and appropriate clones were selected, ligated into plasmid and sequenced.

V Homology Searching of cDNA Clones and Deduced Proteins

Each sequence so obtained was compared to sequences in GenBank using a search algorithm developed by Applied Biosystems Inc. and incorporated into the INHERIT™ 670 Sequence Analysis System. In this algorithm, Pattern Specification Language (developed by TRW Inc., Los Angeles CA) was used to determine regions of homology. The three parameters that determine how the sequence comparisons run were window size, window offset, and error tolerance. Using a combination of these three parameters, the DNA database was searched for sequences

containing regions of homology to the query sequence, and the appropriate sequences were scored with an initial value. Subsequently, these homologous regions were examined using dot matrix homology plots to distinguish regions of homology from chance matches. Smith-Waterman alignments of the protein sequence were used to display the results of the homology search.

Peptide and protein sequence homologies were ascertained using the INHERIT 670 Sequence Analysis System in a way similar to that used in DNA sequence homologies. Pattern Specification Language and parameter windows were used to search protein databases for sequences containing regions of homology which were scored with an initial value. Dot-matrix homology plots were examined to distinguish regions of significant homology from chance matches.

When all three possible predicted translations of the sequence were searched against protein databases such as SwissProt and PIR, no exact matches were found to the possible translations of HPRL. Figure 1 shows the comparison of the HPRL amino acid sequence with the rat PRL-1 amino acid sequence. There is substantial homology among both polypeptides over all the polypeptide sequence with an overall homology of about 90%. The HPRL sequence naturally contains the active site residues, HCXAGXXR, characteristic of this subset of PTPases including the rat PRL-1 PTPase.

Alternatively, BLAST, which stands for Basic Local Alignment Search Tool, was used to search for local sequence alignments (Altschul SF (1993) J Mol Evol 36:290-300; Altschul, SF et al (1990) J Mol Biol 215:403-10). BLAST produces alignments of both nucleotide and amino acid sequences to determine sequence similarity. Because of the local nature of the alignments, BLAST is especially useful in determining exact matches or in identifying homologues. Although it is ideal for matches which do not contain gaps, it is inappropriate for performing motif-style searching. The fundamental unit of BLAST algorithm output is the high-scoring segment pair (HSP).

An HSP consists of two sequence fragments of arbitrary but equal lengths whose alignment is locally maximal and for which the alignment score meets or exceeds a threshold or cutoff score set by the user. The BLAST approach is to look for HSPs between a query sequence and a database sequence, to evaluate the statistical significance of any matches found, and to report only those matches which satisfy the user-selected threshold of significance. The parameter E establishes the statistically significant threshold for reporting database sequence matches. E is interpreted as the upper bound of the expected frequency of chance occurrence of an HSP (or set of HSPs) within the context of the entire database search. Any database sequence whose

match satisfies E is reported in the program output. An E greater than or equal to 25 usually indicates that a match is significant.

BLAST results showed that the polypeptide first identified in the coding sequence of Incyte Clone 3752 of the subject invention had an E parameter value of 144 when compared with polypeptide sequence of the rat prl-1 gene (GenBank accession number L27843). HPRL showed much lower matches with all other polypeptides screened, with the second best matches having E parameter values of 5. No human matches showed up.

VI Antisense analysis

Knowledge of the cDNA sequence of the hprl gene will enable its use in antisense technology in the investigation of gene function. Oligonucleotides, genomic or cDNA fragments comprising the antisense strand of hprl are used either in vitro or in vivo to inhibit expression of the protein. Such technology is now well known in the art, and probes are designed at various locations along the nucleotide sequence. By transfecting cells or whole test animals with such antisense sequences, the gene of interest is effectively be turned off. Frequently, the function of the gene is ascertained by observing behavior at the cellular, tissue or organismal level (e.g. changes in secretory pathways, lethality, loss of differentiated function, changes in morphology, etc.).

In addition to using sequences constructed to interrupt transcription of the open reading frame, modifications of gene expression are obtained by designing antisense sequences to intron regions, promoter/enhancer elements, or even to trans-acting regulatory genes. Similarly, inhibition can be achieved using Hogeboom base-pairing methodology, also known as "triple helix" base pairing.

VII Expression of HPRL

Expression of HPRL is accomplished by subcloning the cDNAs into appropriate expression vectors and transfecting the vectors into appropriate expression hosts. In this particular case, the cloning vector used in the generation of the full length clone also provides for direct expression of the included hprl sequence in E. coli. Upstream of the cloning site, this vector contains a promoter for ß-galactosidase, followed by sequence containing the amino-terminal Met and the subsequent 7 residues of ß-galactosidase. Immediately following these eight residues is an engineered bacteriophage promoter useful for artificial priming and transcription and a number of unique restriction sites, including Eco RI, for cloning.

Induction of the isolated, transfected bacterial strain with IPTG using standard methods produces a fusion protein corresponding to

the first seven residues of ß-galactosidase, about 15 residues of "linker", and the peptide encoded within the cDNA. Since cDNA clone inserts are generated by an essentially random process, there is one chance in three that the included cDNA will lie in the correct frame for proper translation. If the cDNA is not in the proper reading frame, it can be obtained by deletion or insertion of the appropriate number of bases by well known methods including in vitro mutagenesis, digestion with exonuclease III or mung bean nuclease, or oligonucleotide linker inclusion.

The hprl cDNA is shuttled into other vectors known to be useful for expression of protein in specific hosts. Oligonucleotide amplimers containing cloning sites as well as a segment of DNA sufficient to hybridize to stretches at both ends of the target cDNA (25 bases) are synthesized chemically by standard methods. These primers are then used to amplify the desired gene segments by PCR. The resulting new gene segments are digested with appropriate restriction enzymes under standard conditions and isolated by gel electrophoresis. Alternately, similar gene segments are produced by digesting the cDNA with appropriate restriction enzymes and filling in the missing gene segments with chemically synthesized oligonucleotides. Segments of the coding sequence from more than one gene are ligated together and cloned in appropriate vectors to optimize expression of recombinant sequence.

Suitable expression hosts for such chimeric molecules include but are not limited to mammalian cells such as Chinese Hamster Ovary (CHO) and human 293 cells, insect cells such as Sf9 cells, yeast cells such as <u>Saccharomyces</u> <u>cerevisiae</u>, and bacteria such as <u>E</u>. <u>coli</u>. For each of these cell systems, a useful expression vector also includes an origin of replication to allow propagation in bacteria and a selectable marker such as the ß-lactamase antibiotic resistance gene to allow selection in bacteria. In addition, the vectors may include a second selectable marker such as the neomycin phosphotransferase gene to allow selection in transfected eukaryotic host cells. Vectors for use in eukaryotic expression hosts may require RNA processing elements such as 3' polyadenylation sequences if such are not part of the cDNA of interest.

Additionally, the vector may contain promoters or enhancers which increase gene expression. Such promoters are host specific and include MMTV, SV40, or metallothionine promoters for CHO cells; trp, lac, tac or T7 promoters for bacterial hosts, or alpha factor, alcohol oxidase or PGH promoters for yeast. Transcription enhancers, such as the rous sarcoma virus (RSV) enhancer, may be used in mammalian host cells. Once homogeneous cultures of recombinant cells are obtained

through standard culture methods, large quantities of recombinantly produced HPRL can be recovered from the conditioned medium and analyzed using chromatographic methods known in the art.

VIII Isolation of Recombinant HPRL

HPRL is expressed as a chimeric protein with one or more additional polypeptide domains added to facilitate protein purification. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle WA). The inclusion of a cleavable linker sequence such as Factor XA or enterokinase (Invitrogen) between the purification domain and the hprl sequence is useful to facilitate expression of HPRL.

IX Production of HPRL Specific Antibodies

Two approaches are utilized to raise antibodies to HPRL, and each approach is useful for generating either polyclonal or monoclonal antibodies. In one approach, denatured protein from the reverse phase HPLC separation is obtained in quantities up to 75 mg. This denatured protein can be used to immunize mice or rabbits using standard protocols; about 100 micrograms are adequate for immunization of a mouse, while up to 1 mg is used to immunize a rabbit. For identifying mouse hybridomas, the denatured protein is radioiodinated and used to screen potential murine B-cell hybridomas for those which produce antibody. This procedure requires only small quantities of protein, such that 20 mg would be sufficient for labeling and screening of several thousand clones.

In the second approach, the amino acid sequence of HPRL, as deduced from translation of the cDNA, is analyzed to determine regions of high immunogenicity. Oligopeptides comprising appropriate hydrophilic regions, as shown in Figure 2, are synthesized and used in suitable immunization protocols to raise antibodies. Analysis to select appropriate epitopes is described by Ausubel FM et al., supra. The optimal amino acid sequences for immunization are usually at the C-terminus, the N-terminus and those intervening, hydrophilic regions of the polypeptide which are likely to be exposed to the external environment when the protein is in its natural conformation.

Typically, selected peptides, about 15 residues in length, are synthesized using an Applied Biosystems Peptide Synthesizer Model 431A using fmoc-chemistry and coupled to keyhole limpet hemocyanin (KLH, Sigma) by reaction with M-maleimidobenzoyl-N- hydroxysuccinimide ester (MBS; Ausubel FM et al, supra). If necessary, a cysteine may be

introduced at the N-terminus of the peptide to permit coupling to KLH. Rabbits are immunized with the peptide-KLH complex in complete Freund's adjuvant. The resulting antisera are tested for antipeptide activity by binding the peptide to plastic, blocking with 1% BSA, reacting with antisera, washing and reacting with labeled (radioactive or fluorescent), affinity purified, specific goat anti-rabbit IgG.

Hybridomas are prepared and screened using standard techniques. Hybridomas of interest are detected by screening with labeled HPRL to identify those fusions producing the monoclonal antibody with the desired specificity. In a typical protocol, wells of plates (FAST; Becton-Dickinson, Palo Alto CA) are coated with affinity purified, specific rabbit-anti-mouse (or suitable anti-species Ig) antibodies at 10 mg/ml. The coated wells are blocked with 1% BSA, washed and exposed to supernatants from hybridomas. After incubation the wells are exposed to labeled HPRL, 1 mg/ml. Clones producing antibodies will bind a quantity of labeled HPRL which is detectable above background. Such clones are expanded and subjected to 2 cycles of cloning at limiting dilution (1 cell/3 wells). Cloned hybridomas are injected into pristane mice to produce ascites, and monoclonal antibody is purified from mouse ascitic fluid by affinity chromatography on Protein A. Monoclonal antibodies with affinities of at least 10^8 M^{-1}, preferably 10^9 to 10^{10} or stronger, will typically be made by standard procedures as described in Harlow and Lane (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor NY; and in Goding (1986) Monoclonal Antibodies: Principles and Practice, Academic Press, New York City, both incorporated hereby by reference.

X Diagnostic Test Using HPRL Specific Antibodies

Particular HPRL antibodies are useful for the detection or diagnosis of conditions which are characterized by differences in the amount, distribution or interactions of HPRL in various tissues or cells. Specifically, these conditions include those related to abnormal levels of cellular proliferation leading to tumor formation, or to abnormal tissue regeneration responses, such as during liver regeneration.

Diagnostic tests for HPRL include methods utilizing the antibody and a label to detect HPRL in human body fluids, tissues or extracts of such tissues. The polypeptides and antibodies of the present invention are used with or without modification. Frequently, the polypeptides and antibodies will be labeled by joining them, either covalently or noncovalently, with a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and have been reported extensively in both the

scientific and patent literature. Suitable labels include radionuclides, enzymes, substrates, cofactors, inhibitors, fluorescent agents, chemiluminescent agents, chromogenic agents, magnetic particles and the like. Patents teaching the use of such labels include US Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241. Also, recombinant immunoglobulins are produced as shown in US Patent No. 4,816,567, incorporated herein by reference.

A variety of protocols for measuring soluble or membrane-bound HPRL, using either polyclonal or monoclonal antibodies specific for the respective protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA) and fluorescent activated cell sorting (FACS). A two-site monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on HPRL is preferred, but a competitive binding assay may be employed. These assays are described, among other places, in Maddox, DE et al (1983) J Exp Med 158:1211.

XI Purification of Native HPRL Using Specific Antibodies

Native or recombinant HPRL is purified by immunoaffinity chromatography using antibodies specific for HPRL. In general, an immunoaffinity column is constructed by covalently coupling the anti-HPRL antibody to an activated chromatographic resin.

Polyclonal immunoglobulins are prepared from immune sera either by precipitation with ammonium sulfate or by purification on immobilized Protein A (Pharmacia LKB Biotechnology, Piscataway, NJ). Likewise, monoclonal antibodies are prepared from mouse ascites fluid by ammonium sulfate precipitation or chromatography on immobilized Protein A. Partially purified immunoglobulin is covalently attached to a chromatographic resin such as CnBr-activated Sepharose (Pharmacia LKB Biotechnology). The antibody is coupled to the resin, the resin is blocked, and the derivative resin is washed according to the manufacturer's instructions.

Such immunoaffinity columns are utilized in the purification of HPRL by preparing a fraction from cells containing HPRL in a soluble form. This preparation is derived by solubilization of the whole cell or of a subcellular fraction obtained via differential centrifugation by the addition of detergent or by other methods well known in the art. Alternatively, soluble HPRL containing a signal sequence is secreted in useful quantity into the medium in which the cells are grown.

A soluble HPRL-containing preparation is passed over the immunoaffinity column, and the column is washed under conditions that

allow the preferential absorbance of HPRL proteins. Then, the column is eluted under conditions that disrupt antibody/HPRL binding (e.g., a buffer of pH 2-3 or a high concentration of a chaotrope such as urea or thiocyanate ion), and HPRL is collected.

XII HPRL Cellular Localization and Function

HPRL's localization in the nucleus of mast cells is confirmed in the following manner. Cells are lysed by freeze-thaw methods followed by staining of the lysed cell with a radiolabeled monoclonal antibody specific for HPRL.

Once HPRL is localized to a specific cellular compartment, reconstitution studies are performed to investigate its function and its interactions with other cellular components. For example, if HPRL is localized in the nucleus HPRL may act to dephosphorylate specific DNA-binding proteins indirectly affecting the transcription of specific genes that affect cellular proliferation or tissue regeneration. Preferably, these reconstitution studies are performed in cell free systems depleted of naturally occurring HPRL to study the effect of HPRL absence on cells. Then, the concentration of HPRL is gradually increased to recover HPRL activity that affects cellular proliferation and tissue regeneration. This method is used to test HPRL derivatives or other homologs for biological activity.

XIII Drug Screening

HPRL is used to screen compounds that control cellular proliferation. Alternatively, HPRL is used to screen compounds that enhance tissue regeneration. One method of compound screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant nucleic acids expressing the HPRL polypeptide or a fragment thereof. Compounds are screened against such transformed cells by measuring, for example, alterations in HPRL activity by using specific substrates that will be dephosphorylated by HPRL. Preferably, these experiments are performed in cell-free systems so that the screened compounds and phosphorylated HPRL substrates are more accessible to HPRL.

Additionally, the present invention provides <u>in vitro</u> methods of screening for test compounds which affect HPRL activity. These methods comprise contacting a compound with HPRL and assaying for the presence of a complex between the compound and HPRL by methods well known in the art. After suitable incubation, free compound is separated from that in bound form, and the complexed HPRL is assayed for activity.

Examples of test compounds include those that act to inhibit HPRL, such as antibodies or inhibitors, which may be used to control high levels of cellular proliferation. Alternatively, test compounds

include those that act to increase HPRL activity, such as agonists, which may be used to enhance tissue regeneration processes.

Another technique for drug screening that provides high throughput screening for compounds having suitable binding affinity to HPRL, is described in European Patent 84/03564, incorporated herein by reference.

This invention also contemplates the use of competitive drug screening assays in which neutralizing antibodies capable of binding HPRL specifically compete with a test compound for binding to HPRL. In this manner, the antibodies are used to detect the presence of any peptide which shares one or more antigenic determinants with HPRL.

XIV Rational Drug Design

The goal of rational drug design is to produce structural analogs of biologically active polypeptides of interest or of small molecules with which they interact, including nonhydrolyzable analogs of phosphorylated substrates for HPRL. Any of these examples can be used to fashion drugs which are more active or stable forms of the polypeptide or which enhance or interfere with the function of a polypeptide in vivo (Hodgson J (1991) Bio/Technology 9:19-21, incorporated herein by reference).

In one approach, the three-dimensional structure of a protein of interest, or of a protein-inhibitor complex, is determined by x-ray crystallography, by computer modeling or, most typically, by a combination of the two approaches. Both the shape and charges of the polypeptide are ascertained to elucidate the structure and to determine active site(s) of the molecule. Less often, useful information regarding the structure of a polypeptide is gained by modeling based on the structure of homologous proteins. In both cases, relevant structural information is used to design analogous HPRL-like molecules or to identify efficient inhibitors. Useful examples of rational drug design include molecules which have improved activity or stability as shown by Braxton S and Wells JA (1992) Biochemistry 31:7796-7801 or which act as inhibitors, agonists, or antagonists of native peptides as shown by Athauda SB et al (1993) J Biochem 113:742-746, incorporated herein by reference.

It is also possible to isolate a target-specific antibody, selected by functional assay, as described above, and then to solve its crystal structure. This approach, in principle, yields a pharmacore upon which subsequent drug design can be based. It is possible to bypass protein crystallography altogether by generating anti-idiotypic antibodies (anti-ids) to a functional, pharmacologically active antibody. As a mirror image of a mirror image, the binding site of the anti-ids would be expected to be an

analog of the original receptor. The anti-id is then used to identify and isolate peptides from banks of chemically or biologically produced peptides. The isolated peptides would then act as the pharmacore.

By virtue of the present invention, sufficient amount of polypeptide is made available to perform such analytical studies as X-ray crystallography. In addition, knowledge of the HPRL amino acid sequence provided herein will provide guidance to those employing computer modeling techniques in place of or in addition to x-ray crystallography.

XV Use and Administration of Drugs

HPRL appears to function in cellular proliferation and tissue regeneration. Antibodies against HPRL or compounds that inhibit HPRL activity, may be used to treat abnormal cellular proliferation. HPRL, fragments or derivatives thereof and HPRL agonists may be used to enhance tissue regeneration. All these compounds with potential therapeutic effects are termed therapeutic compounds.

Therapeutic compounds will be formulated in a nontoxic, inert, pharmaceutically acceptable aqueous carrier medium preferably at a pH of about 5 to 8, more preferably 6 to 8, although the pH may vary according to the characteristics of the formulation and its administration. Characteristics such as solubility of the molecule, half-life and antigenicity/immunogenicity will aid in defining an effective carrier. Recombinant, organic or synthetic molecules resulting from drug design may be equally effective in particular situations.

Therapeutic compounds may be delivered by known routes of administration including but not limited to topical creams and gels; transmucosal spray and aerosol, transdermal patch and bandage; injectable, intravenous and lavage formulations; and orally administered liquids and pills, particularly formulated to resist stomach acid and enzymes. The particular formulation, exact dosage, and route of administration will be determined by the attending physician and will vary according to each specific situation.

Such determinations are made by considering multiple variables such as the condition to be treated, the therapeutic compound to be administered, and the pharmacokinetic profile of the particular therapeutic compound. Additional factors which may be taken into account include disease state (e.g. severity) of the patient, age, weight, gender, diet, time of administration, drug combination, reaction sensitivities, and tolerance/response to therapy. Long acting therapeutic compound formulations might be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular therapeutic HPRL.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature; see US Patent Nos. 4,657,760; 5,206,344; and 5,225,212. It is anticipated that different formulations will be effective for different uses of therapeutic compounds and that administration targeting a tissue or organ may necessitate delivery in a specific manner.

All publications and patents mentioned in the above specification are herein incorporated by reference. The foregoing written specification is considered to be sufficient to enable one skilled in the art to practice the invention. Indeed, various modifications of the above described modes for carrying out the invention which are readily apparent to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: INCYTE PHARMACEUTICALS, INC.

(ii) TITLE OF INVENTION: PROTEIN TYROSINE PHOSPHATASE DERIVED FROM HUMAN MAST CELLS

(iii) NUMBER OF SEQUENCES: 2

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
(B) STREET: 3174 Porter Drive
(C) CITY: Palo Alto
(D) STATE: CA
(E) COUNTRY: USA
(F) ZIP: 94304

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) PCT APPLICATION NUMBER:PCT/US96/12665
(B) FILING DATE: 01-AUG-1996
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION SERIAL NUMBER:US 08/567,507
(B) FILING DATE: 05-DEC-1995

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION SERIAL NUMBER: US 60/002,151
(B) FILING DATE: 10-AUG-1995

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Billings, Lucy J.
(B) REGISTRATION NUMBER: 36,749
(C) REFERENCE/DOCKET NUMBER: PF-0047 PCT

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 415-855-0555
(B) TELEFAX: 415-845-4166

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1144 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: Human Mast Cells
(B) CLONE: 3752

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
TTTGGGTTCA AGTGGATTCT AAATACTTTG CTTATCTTGA AGAGAGAAGC TTCATAAGGA     60
ATAAACAAGT TGAATAGAGA AAACACTGAT TGATAATAGG CATTTTAGTG GTCTTTTTAA    120
TGTTTTCTGC TGTGAAACAT TTCAAGATTT ATTGATTTTT TTTTTTCACT TTCCCCATCA    180
CACTCACACG CACGCTCACA CTTTTTATTT GCCATAATGA ACCGTCCAGC CCCTGTGGAG    240
ATCTCCTATG AGAACATGCG TTTTCTGATA ACTCACAACC CTACCAATGC TACTCTCAAC    300
AAGTTCACAG AGGAACTTAA GAAGTATGGA GTGACGACTT TGGTTCGAGT TTGTGATGCT    360
ACATATGATA AAGCTCCAGT TGAAAAAGAA GGAATCCACG TCCTAGATTG GCCATTTGAT    420
GATGGAGCTC CACCCCCTAA TCAGATAGTA GATGATTGGT TAAACCTGTT AAAAACCAAA    480
TTTCGTGAAG AGCCAGGTTG CTGTGTTGCA GTGCATTGTG TTGCAGGATT GGGAAGGGCA    540
CCTGTGCTGG TTGCACTTGC TTTGATTGAA TGTGGAATGA AGTACGAAGA TGCAGTTCAG    600
TTTATAAGAC AAAAAAGAAG GGGAGCGTTC AATTCCAAAC AGCTGCTTTA TTTGGAGAAA    660
TACCGACCTA AGATGCGATT ACGCTTCAGA GATACCAATG GGCATTGCTG TGTTCAGTAG    720
AAGGAAATGT AAACGAAGGC TGACTTGATT GTGCCATTTA GAGGGAACTC TTGGTACCTG    780
GAAATGTGAA TCTGGAATAT TACCTGTGTC ATCAAAGTAG TGATGGATTC AGTACTCCTC    840
AACCACTCTC CTAATGATTG GAACAAAAGC AAACAAAAAA GAAATCTCTC TATAAAATGA    900
ATAAAATGTT TAAGAAAAGA GAAAGAGAAA AGGAATTANT CAGTGAAGGG TGATTTTGGT    960
CCTAGTTTGG GAGTTTGATT TCTGNCAGGA TTGAATTATT TGGAAATCTC CTGTCTTTTT   1020
AAACTTTTTC GTNAATAGGG TCTCTTAAGG GAAAACCAGC AGAACATTAG CCTGTGCAAA   1080
ACCACTTTTT GGGGGAGAAA CTCTTCCATT ATGTTGGGAN ATAGACTCCC TGGGNGGGGA   1140
TTTT                                                                1144
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 167 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Asn Arg Pro Ala Pro Val Glu Ile Ser Tyr Glu Asn Met Arg Phe
  1               5                  10                  15

Leu Ile Thr His Asn Pro Thr Asn Ala Thr Leu Asn Lys Phe Thr Glu
             20                  25                  30
```

```
Glu Leu Lys Lys Tyr Gly Val Thr Thr Leu Val Arg Val Cys Asp Ala
         35                  40                  45

Thr Tyr Asp Lys Ala Pro Val Glu Lys Glu Gly Ile His Val Leu Asp
     50                  55                  60

Trp Pro Phe Asp Asp Gly Ala Pro Pro Pro Asn Gln Ile Val Asp Asp
 65                  70                  75                  80

Trp Leu Asn Leu Leu Lys Thr Lys Phe Arg Glu Glu Pro Gly Cys Cys
                 85                  90                  95

Val Ala Val His Cys Val Ala Gly Leu Gly Arg Ala Pro Val Leu Val
            100                 105                 110

Ala Leu Ala Leu Ile Glu Cys Gly Met Lys Tyr Glu Asp Ala Val Gln
        115                 120                 125

Phe Ile Arg Gln Lys Arg Arg Gly Ala Phe Asn Ser Lys Gln Leu Leu
    130                 135                 140

Tyr Leu Glu Lys Tyr Arg Pro Lys Met Arg Leu Arg Phe Arg Asp Thr
145                 150                 155                 160

Asn Gly His Cys Cys Val Gln
                165
```

CLAIMS

1. A purified polynucleotide comprising a nucleic acid sequence encoding the polypeptide of SEQ ID NO:2, or the complement of said polynucleotide.

2. The polynucleotide of Claim 1 comprising the nucleic acid sequence for the human homolog rat PRL-1 (hprl) of SEQ ID NO:1.

3. An antisense molecule comprising the complement of the polynucleotide of Claim 2 or a portion thereof.

4. A pharmaceutical composition comprising the antisense molecule of Claim 3 and a pharmaceutically acceptable excipient.

5. A method of treating a subject with a condition associated with altered hprl expression comprising administering an effective amount of the pharmaceutical composition of Claim 4 to the subject.

6. A diagnostic composition comprising an oligomer of the polynucleotide of Claim 2.

7. A diagnostic test for a condition associated with altered hprl expression comprising the steps of:

a) providing a biological sample;

b) combining the biological sample and the diagnostic composition of Claim 6;

c) allowing hybridization to occur between the biological sample and the diagnostic composition under suitable conditions;

d) measuring the amount of hybridization to obtain a sample value; and

e) comparing the sample value with standard values to determine whether hprl expression is altered.

8. An expression vector comprising the polynucleotide of Claim 1.

9. A host cell transformed with the expression vector of Claim 8.

10. A method for producing a polypeptide, said method comprising the steps of:

a) culturing the host cell of Claim 9 under conditions suitable for the expression of the polypeptide; and

b) recovering the polypeptide from the host cell culture.

11. A purified polypeptide (HPRL) comprising the amino acid sequence of SEQ ID NO:2.

12. A diagnostic composition comprising the polypeptide of Claim 11 or a portion thereof.

13. A pharmaceutical composition comprising the polypeptide of Claim 11 and a pharmaceutically acceptable excipient.

14. A method of treating a subject with a condition associated with altered HPRL expression comprising administering an effective amount of the pharmaceutical composition of Claim 13 to the subject.

15. An antibody specific for the purified polypeptide of Claim 11, or

portion thereof.

16. A diagnostic composition comprising the antibody of Claim 15.

17. A diagnostic test for a condition associated with altered HPRL expression comprising the steps of:

a) providing a biological sample;

b) combining the biological sample and the antibody of Claim 15 under conditions suitable for complex formation;

c) measuring the amount of complex formation between HPRL and the antibody to obtain a sample amount; and

d) comparing the amount of complex formation in the sample with standard amounts of complex formation, wherein a variation between the sample amount and standard amounts of complex formation establishes the presence of the condition.

18. A method of screening a plurality of compounds for specific binding affinity with the polypeptide of Claim 11 or any portion thereof comprising the steps of:

a) providing a plurality of compounds;

b) combining HPRL with each of a plurality of compounds for a time sufficient to allow binding under suitable conditions; and

c) detecting binding of HPRL to each of the plurality of compounds, thereby identifying the compounds which specifically bind HPRL.

```
    . . . M N R P A P V E . . Y . N M R F L I T H N P T N A T L
    M A R M N R P A P V E V S Y E N M R F L I T H N P T N A T L
                      10                            20                            30
1   - - - M N R P A P V E I S Y E N M R F L I T H N P T N A T L
1   M A R M N R P A P V E V T Y K N M R F L I T H N P T N A T L

    N K F . E E L K K Y G V T T . V R V C . A T Y D . . . V E K
    N K F I E E L K K Y G V T T L V R V C D A T Y D T A L V E K
                      40                            50                            60
28  N K F T E E L K K Y G V T T L V R V C D A T Y D K A P V E K
31  N K F I E E L K K Y G V T T I V R V C E A T Y D T T L V E K

    E G I H V L D W P F D D G A P P . N Q I V D D W L . L . K .
    E G I H V L D W P F D D G A P P S N Q I V D D W L S L V K I
                      70                            80                            90
58  E G I H V L D W P F D D G A P P P N Q I V D D W L N L L K T
61  E G I H V L D W P F D D G A P P S N Q I V D D W L S L V K I

    K F R E E P G C C . A V H C V A G L G R A P V L V A L A L I
    K F R E E P G C C V A V H C V A G L G R A P V L V A L A L I
                      100                           110                           120
88  K F R E E P G C C V A V H C V A G L G R A P V L V A L A L I
91  K F R E E P G C C I A V H C V A G L G R A P V L V A L A L I

    E . G M K Y E D A V Q F I R Q K R R G A F N S K Q L L Y L E
    E G G M K Y E D A V Q F I R Q K R R G A F N S K Q L L Y L E
                      130                           140                           150
118 E C G M K Y E D A V Q F I R Q K R R G A F N S K Q L L Y L E
121 E G G M K Y E D A V Q F I R Q K R R G A F N S K Q L L Y L E

    K Y R P K M R L R F . D . N G H . . . C C . Q -
    K Y R P K M R L R F K D S N G H R N N C C V Q -
                      160                           170
148 K Y R P K M R L R F R D T N G H - - - C C V Q .
151 K Y R P K M R L R F K D S N G H R N N C C I Q .
```

FIGURE 1

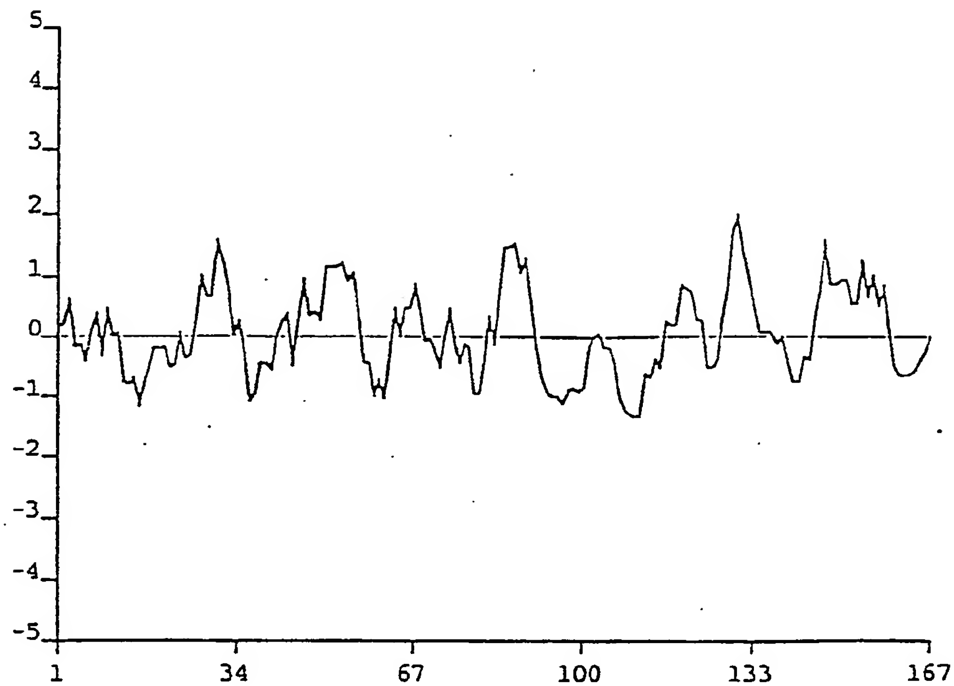

FIGURE 2

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 96/12665

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/55 //C12N9/16

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C07K G01N C12P C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HUMAN GENETICS, vol. 96, 1995, DE, pages 532-538, XP000196605 M. MONTAGNA ET AL.: "A 100-kb physical and transcriptional map around the EDH17B2 gene: identification of three novel genes and a pseudogene of a human homologue of the rat PRL-1 tyrosine phosphatase" see figure 4A	1-3
Y		6-13, 15-17
A		18
X	& DATABASE EMBL/GENBANK/DDBJ Accession number U14603, 10 July 1995 G.M. LENOIR: "HUMAN PROTEIN-TYROSINE PHOSPHATASE (HU-PP-1) mRNA, partial sequence"	1-3

	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

2

Date of the actual completion of the international search	Date of mailing of the international search report
28 November 1996	17. 12. 96

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Mateo Rosell, A.M.

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	MOLECULAR CELLULAR BIOLOGY, vol. 14, no. 6, 1994, USA, pages 3752-3762, XP000196604 R.H. DIAMOND ET AL.,: "PRL-1, a unique nuclear protein tyrosine phosphatase, affects cell growth" cited in the application see figure 2C; table 1	11-13
Y	see figures 5,6	15-17
Y	WO,A,91 13989 (WASHINGTON RES FOUND) 19 September 1991 see claims	8-10
Y	CANCER RESEARCH, vol. 53, no. 10, 15 May 1993, USA, pages 2272-2278, XP000609132 Y.F. ZHAI ET AL.: "Increased expression of specific protein tyrosine phosphatases in human breast epithelial cells neoplastically transformed by the neu oncogene" figures 1 and 6, specially see the whole document	6,7, 15-17
A	WO,A,95 06735 (LUDWIG INSTITUTE FOR CANCER RESEARCH) 9 March 1995 see the whole document	3,4, 15-17
A	WO,A,94 24161 (UNIV NEW YORK ; MEDICAL CENTER) 27 October 1994 see the whole document	6,7, 15-17
P,X	GENOMICS, vol. 28, no. 3, 10 August 1995, USA, pages 530-542, XP000196618 J.M. ROMMENS ET AL.: "Generation of a transcription map at the HSD17B locus centromeric to BRCA1 at 17q21" See results: identification of human homologues of two rat genes, p534-535. see table 1	2
P,X	GENOMICS, vol. 35, no. 1, 1 June 1996, USA, pages 172-181, XP000196619 Z. ZHAO ET AL., : "Characterization and genomic mapping of genes and pseudogenes of a new human protein tyrosine phosphatase" see the whole document, specially figure 1A	2

2

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 96/12665

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
   because they relate to subject matter not required to be searched by this Authority, namely:
   Remark : Incomplete search : Claim 5, 14: fail to comply with Rule 39(iv)

   From PCT (subject matter under article 17 2(a) and (b)

2. [ ] Claims Nos.:
   because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
   because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searches without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 96/12665

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9113989	19-09-91	AT-T- 123064	15-06-95
		CA-A- 2078010	15-09-92
		DE-D- 69110034	29-06-95
		DE-T- 69110034	05-10-95
		EP-A- 0520029	30-12-92
		EP-A- 0627489	07-12-94
		ES-T- 2073165	01-08-95
WO-A-9506735	09-03-95	AU-A- 7644394	22-03-95
		CA-A- 2170515	09-03-95
WO-A-9424161	27-10-94	AU-A- 6710094	08-11-94